

Contrôle continu : Bioanalyse (EL6BIOFM) – 6 mars 2012

Question 1

Si dans un génome, des gènes appartiennent à une famille multigénique, nous dirons qu'ils sont : 1) homologues, 2) orthologues, 3) paralogues ? Argumenter votre (vos) choix.

Question 2

- a) Expliquer la différence entre les banques de données EMBL et TrEMBL
- b) Expliquer ce que sont les matrices de substitutions. Pourquoi sont elles identifiées par des numéros différents (ex : PAM120 et PAM350 ou BLOSUM62 et BLOSUM30).
- c) Pour effectuer quel(s) type(s) d'analyse(s) utiliseriez-vous les programmes suivants :
 - 1) SRS, 2) dotpath de la suite EMBOSS, et 3) strecher de la suite EMBOSS.

Question 3

a) Vous souhaitez aligner deux séquences d'acides nucléiques, l'une ayant une taille de 5600 pb et l'autre de 1300 pb. Vous disposez de deux programmes, l'un réalisant un alignement global et l'autre un alignement local. Lequel de ces deux programmes utiliseriez-vous? Justifier votre réponse.

b) Utilisez la méthode de programmation dynamique pour déterminer l'alignement local optimal entre les deux séquences suivantes :

Séquence 1 : AGTCATG

Séquence 2 : TGATA

Système de scores : identité = 2, substitution = -1, indel = -2,5 (Utilisation pour le calcul d'un score d'homologie)

Remplir la matrice de programmation dynamique et produire l'alignement final. Quel est le score de cet alignement ? Comment l'avez-vous obtenu?

c) Expliquer pourquoi la pénalité des indels doit être plus importante que la pénalité des substitutions.

Question 4

La requête de la Figure 1 a été effectuée à l'aide du logiciel SRS.

- a) Quelle base de données a été interrogée ?
- b) Quelle question était posée à la base de données?
- c) Cette requête vous a permis d'obtenir la fiche fournie en Annexe 1 dont le champ séquence a été supprimé. Répondez aux questions suivantes :
 - quelle est la taille de la séquence ?
 - quelles sont les positions des introns ?
 - la protéine codée par ce gène est-elle disponible dans une base de données ? Si oui laquelle et quel est son numéro d'accèsion dans cette base?
 - le lotus est-il une monocotylédone ? Où avez-vous trouvé l'information.

Figure 1

Standard Query Form

EMBL-EBI Enter Text Here Find Help Feedback

Databases Tools Research Training Industry About Us Help Site Index RSS

Quick Search Library Page Query Form Tools Results Projects Views Databanks HELP

Reset search EMBL

Search Options

Combine search terms with: & (AND)

Use wildcards

Get results of type: Entry

Fields you can search

Fields you can search	Your search terms
Description	Leghemoglobin
Organism Name	lotus japonicus
Description	complete cds
AllText	

In a single field, you can separate multiple values by: &, | or ! Search

Result Display Options

View results using: EMBLSeqSimpleView

or

Create a view

Show 30 results per page

Create a view

Select the fields you want displayed in your view and choose the format

Choose 1 or more fields: ID, Topology, Molecule, Data Class, Division, Sequence Length, Accession Number

Display As: Table List

Sequence Format: embi

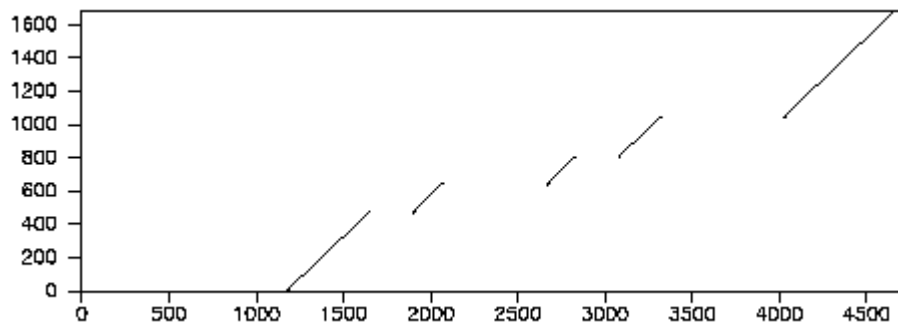
Search

Tips

To do more advanced queries, use the [Extended Query](#) Form.

Question 5

La matrice de point (dotplot) suivante a été obtenue en comparant les séquences d'un gène et de son ARN messager.



- Sur quel axe se trouve la séquence correspondant au gène ? Donnez un ordre d'idée de sa longueur
- Que représentent les traits obliques : 1) en terme d'analyse bioinformatique, 2) en terme biologique

Question 6

Vous avez réalisé l'alignement suivant avec le programme stretch de la suite EMBOSS.

- Quelle matrice de substitution a été utilisée ? Quels sont les pondérations utilisées pour les indels (appelés gaps en anglais). Expliquer leur différence.
- Expliquer à quoi correspondent les différents pourcentages obtenus.
- Expliquer ce qui est représenté sur la ligne intermédiaire de l'alignement (ligne entre les deux séquences).
- Quelles sont la(les) région(s) de ces deux protéines qui sont les mieux conservées.

Aligned_sequences: 2	Length: 334
1: RBSR_ECOLI	Identity: 101/334 (30.2%)
2: RBSR_BACSU	Similarity: 159/334 (47.6%)
Matrix: EBLOSUM62	Gaps: 12/334 (3.6%)
Gap_penalty: 12	Score: 345
Extend_penalty: 2	

```
RBSR_E MATMKDVARLAGVSTSTVSHVINKDRFVSEAITAKVEAAIKELNYAPSAL
      :::::::::: :::: :::: :. . . : . : : : : . : : : : : . .
RBSR_B MATIKDVAGAAGVSVATVSRNLNDNGYVHEETRTRVIAAMAKLNYPNEV
      10      20      30      40      50

      60      70      80      90      100
RBSR_E ARSLKLNQTHITIGMLITASTNPFYSELVRGBVERSCFERGYSLVLCNTEGD
      :::: . . :::: : : : : : : : : : : : : : : . . . .
RBSR_B ARSLYKRESRLIGLLLDPDITNPFPPQLARGAEDELNREGYRLIFGNSDEE
      60      70      80      90      100

      110      120      130      140      150
RBSR_E EQRMNRNLETLMQKRDVGLLLCTETHQPSREIMQRYPTVPTVMMDWAPF
      . : : : : : : : : : : : : : : : : : : : : : : : : :
RBSR_B LKKELEYLQTFKQNHVAGII---AATNYPDLEEYSGM-NYPVVFLD-RTL
      110      120      130      140

      160      170      180      190      200
RBSR_E DGSDSLIQDNSLLGGDLATQYLIDKGHTRIACITGPLDKTPARLRLEGYR
      . : : : : : : : : : : : : : : : : : : : : : : : :
RBSR_B EG-APSVSSDGYTGVKLAAQAIHIGKSQRITLLRGPAPHLPTAQDRFNGAL
      150      160      170      180      190

      210      220      230      240      250
RBSR_E AAMKRAGLNIPDGYEVTGDFEFNGGFDAMRQLLSHPLRPQAVFTGNDAMA
      . : : . . : : : : : : : : : : : : : : : : : : :
RBSR_B EILKQAEVDFQ--VIETASFSIKDAQSMAKELFASYPATDGVIASNDIQA
      200      210      220      230      240

      260      270      280      290      300
RBSR_E VGVYQALYQAELQVPQDIAVIGYDDIELASFMTPLTTIHQPKDELGELA
      : . . : : : : : : : . . : : : : : : : : . : . :
RBSR_B AAVLHEALRRGKNVPEDIQIIGYDDIPQSGLLFPPLSTIKQPAYDMGKEA
      250      260      270      280      290

      310      320      330
RBSR_E IDVLIHRITQPTLQQQLQLTPILMERGSA----
      . : . : . : . : . : . : . : . : . : . : . : . :
RBSR_B AKLLLGIKKQPLAETAIQMPVTYIGRKTRTKED
      300      310      320
```

```

ID  AB042718; SV 1; linear; genomic DNA; STD; PLN; 6653 BP.
XX
AC  AB042718;
XX
DT  18-NOV-2000 (Rel. 65, Created)
DT  20-APR-2006 (Rel. 87, Last updated, Version 4)
XX
DE  Lotus japonicus lljlb gene for leghemoglobin, complete cds.
XX
KW  .
XX
OS  Lotus japonicus
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC  rosids;fabids; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.
XX
RN  [2]
RA  Uchiumi T., Tsuruta T., Suzuki A., Abe M., Higashi S.;
RT  "Genomic leghemoglobin genes of Lotus japonicus";
RL  Unpublished.
XX
FH  Key                Location/Qualifiers
FH
FT  source              1..6653
FT                      /organism="Lotus japonicus"
FT                      /ecotype="Gifu"
FT                      /mol_type="genomic DNA"
FT                      /note="synonym: Lotus corniculatus var. japonicus"
FT                      /db_xref="taxon:34305"
FT  mRNA                join(3055..3371,3494..3602,3728..3835,4638..5357)
FT  CDS                 join(3277..3371,3494..3602,3728..3835,4638..4769)
FT                      /codon_start=1
FT                      /gene="lljlb"
FT                      /product="leghemoglobin"
FT                      /db_xref="GOA:Q9FEP8"
FT                      /db_xref="InterPro:IPR000971"
FT                      /db_xref="InterPro:IPR001032"
FT                      /db_xref="InterPro:IPR009050"
FT                      /db_xref="InterPro:IPR012292"
FT                      /db_xref="InterPro:IPR019824"
FT                      /db_xref="UniProtKB/Swiss-Prot:Q9FEP8"
FT                      /protein_id="BAB18108.1"
XX

```